MATHunesa
Jurnal Ilmiah Matematika
e-ISSN : 2716-506X | p-ISSN : 2301-9115

Volume 14 No 02
Tahun 2026

INTEGRASI *MACHINE LEARNING* DENGAN *GEOGRAPHICALLY WEIGHTED REGRESSION* UNTUK ANALISIS SPASIAL INDEKS PEMBANGUNAN MANUSIA

Ismi Rizqa Lina
Program Studi Sains Data, Universitas Insan Cita Indonesia, email : irizqalina@gmail.com*

Dia Cahya Wati
Program Studi Sains Data, Universitas Insan Cita Indonesia, email : diacahyawati@gmail.com

Ibnu Mansyur Hamdani
Program Studi Teknik Perawatan Mesin, Akademi Komunitas Industri Manufaktur Bantaeng, email : ibnumansyur27@gmail.com

Yulia Resti
Program Studi Matematika, Universitas Sriwijaya, email : Yulia_resti@mipa.unsri.ac.id

Abstrak

Keberhasilan pembangunan nasional tidak semata-mata ditentukan oleh pertumbuhan ekonomi tetapi juga oleh kualitas pembangunan manusia, yang diukur dengan Indeks Pembangunan Manusia (IPM). Meskipun IPM Indonesia telah membaik secara nasional, disparitas antarwilayah, terutama di Indonesia bagian barat, tetap menjadi tantangan yang signifikan. Untuk mengatasi hal ini, diperlukan analisis yang komprehensif untuk memahami distribusi spasial IPM dan faktor-faktor yang memengaruhinya. Studi ini mengusulkan pendekatan terpadu yang menggabungkan teknik Pembelajaran Mesin, yaitu *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) dengan *Geographically Weighted Regression* (GWR) menggunakan pembobotan Gaussian adaptif untuk mengklasifikasikan dan menganalisis IPM secara spasial. Algoritma KNN mencapai akurasi klasifikasi tertinggi sebesar 94,23%, sedikit mengungguli SVM pada 92,30%. Analisis GWR berhasil mengelompokkan wilayah studi menjadi 16 klaster spasial dengan faktor-faktor dominan yang berbeda, seperti jumlah penduduk (JP), indeks pengembangan literasi manusia (IPLM), pengeluaran per kapita (PPK), tingkat pengangguran terbuka (TPT), dan jumlah distrik (JK). Sebagai contoh, model GWR di Jakarta Selatan mengidentifikasi pengaruh lokal yang signifikan dari variabel-variabel ini, dengan persamaan model spesifik yang menunjukkan ketergantungan spasial yang kuat. Koefisien determinasi model (R^2) mencapai 84,17%, menunjukkan daya penjelasan yang tinggi. Temuan ini menunjukkan bahwa integrasi Pembelajaran Mesin dan GWR tidak hanya meningkatkan akurasi klasifikasi tetapi juga memberikan wawasan spasial yang lebih mendalam, menawarkan dasar yang kuat untuk perencanaan kebijakan berbasis data yang bertujuan mengurangi disparitas regional dalam IPM.

Kata Kunci: : Indeks Pembangunan Manusia, *Machine Learning, K-Nearest Neighbor, Support Vector Machine, Geographically Weighted Regression*

Abstract

The success of national development is not solely determined by economic growth but also by the quality of human development, which is measured by the Human Development Index (HDI). Although Indonesia's HDI has improved nationally, disparities between regions, particularly in western Indonesiaremain a significant challenge. To address this, a comprehensive analysis is needed to understand the spatial distribution of HDI and its influencing factors. This study proposes an integrated approach that combines Machine Learning techniques, namely K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) with Geographically Weighted Regression (GWR) using adaptive Gaussian weighting to classify and analyze HDI spatially. The KNN algorithm achieved the highest classification accuracy of 94.23%, slightly outperforming SVM at 92.30%. The GWR analysis successfully grouped the study area into 16 spatial clusters with distinct dominant factors, such as population size (JP), human literacy development index (IPLM), per capita expenditure (PPK), open unemployment rate (TPT), and number of districts (JK). For example, the GWR model in South Jakarta identified significant local influences from these variables, with a specific model equation indicating strong spatial dependency. The model's coefficient of determination (R^2) reached 84.17%, indicating high explanatory power. These findings demonstrate that the integration of Machine Learning and GWR not only enhances

classification accuracy but also provides deeper spatial insights, offering a robust basis for data-driven policy planning aimed at reducing regional disparities in HDI.

Keywords: *Human Development Index; Machine Learning; K-Nearest Neighbor; Support Vector Machine; Geographically Weighted Regression.*

PENDAHULUAN

Keberhasilan pembangunan nasional tidak hanya diukur dari tingginya laju pertumbuhan ekonomi, tetapi pencapaian dalam pembangunan manusia. Indeks Pembangunan Manusia (IPM) digunakan sebagai indikator untuk mengamati dan mengevaluasi tingkat pembangunan manusia (Pamungkas and Widiyanto 2022) Meskipun secara nasional IPM Indonesia terus mengalami peningkatan, disparitas antar wilayah masih menjadi tantangan utama, terutama di wilayah Indonesia bagian barat. Disparitas pembangunan manusia di wilayah barat tahun 2013 sebesar 7,66 poin. Rentang disparitas tersebut sudah lebih kecil jika dibandingkan dengan tahun 2011 yaitu 8,32 poin (Indonesia 2017). Oleh karena itu, klasifikasi IPM dan analisis spasial yang lebih mendalam diperlukan untuk memahami faktor-faktor yang memengaruhi distribusi IPM secara lebih akurat.

Salah satu metode untuk mengevaluasi IPM, yaitu dengan mengelompokkan data berdasarkan kategori yang telah ditetapkan. Seiring dengan perkembangan teknologi, proses pengelompokan kini dapat dilakukan dengan bantuan kecerdasan buatan melalui *machine learning* (ML). Di antara metode *machine learning* yang dapat digunakan, yaitu *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM), yang telah terbukti mampu melakukan klasifikasi secara efektif menggunakan algoritma yang sederhana namun tetap akurat dan efisien (Latuconsina, van Delsen, and others 2024) dan (Jun 2021).

KNN bekerja dengan mengelompokkan data berdasarkan kedekatan antar titik, sedangkan SVM mencari *hyperplane* optimal yang memisahkan data dalam ruang multidimensi (Bishop and Nasrabadi 2006). Penelitian tentang klasifikasi IPM menggunakan metode KNN telah dilakukan oleh (Oktaviana, Lubis, and Widyasari 2024), (Safitri, Satria, and Badri 2024), dan (Darsyah 2017). Di sisi lain klasifikasi IPM juga dilakukan dengan metode SVM oleh (Pamungkas and Widiyanto 2022) dan (Latuconsina et al. 2024). Kemudian (Fauzi 2017) dan (Witata and Triloka 2023) menggabungkan metode KNN dan SVM untuk klasifikasi IPM dan prediksi pengangguran di provinsi Lampung. Meskipun *machine learning* memiliki keunggulan dalam mengenali pola dan tren, pendekatan *machine learning* sering kali mengabaikan faktor geografis dalam analisis pembangunan manusia.

Untuk mengatasi keterbatasan tersebut, *Geographically Weighted Regression* (GWR) dapat digunakan sebagai metode pelengkap dalam analisis spasial (Krismayanto and Pasaribu 2022). Pemodelan IPM dengan GWR juga dilakukan oleh (Marizal and Atiqah 2022) dan (Maulana, Meilawati, and Widiastuti 2019). Pengembangan algoritma baru KNN-GWR untuk skala besar dilakukan (Yang et al. 2023), selanjutnya machine-learning-GWR digunakan untuk evaluasi spasial IPM dibagian barat, Adapun algoritma *machine* yang digunakan dalam penelitiannya, yaitu regresi linier berganda dan model regresi ridge (Firmansyah et al. 2023).

Berdasarkan latar belakang di atas, penelitian ini menawarkan pendekatan baru dengan mengintegrasikan metode KNN, SVM, dan GWR untuk klasifikasi dan analisis spasial IPM yang belum dilakukan dalam penelitian sebelumnya. Pendekatan ini mengatasi keterbatasan *machine learning* konvensional yang sering mengabaikan faktor geografis dalam analisis pembangunan manusia. Dengan KNN, klasifikasi dilakukan berdasarkan kedekatan antar data (Darsyah 2017), sementara SVM menentukan *hyperplane* optimal untuk pemisahan kategori (Pamungkas and Widiyanto 2022), dan GWR memperhitungkan heterogenitas spasial dalam distribusi IPM di Indonesia bagian barat (Marizal and Atiqah 2022). Berbeda dengan (Firmansyah et al. 2023) yang menggunakan empat variabel independen, pada penelitian ini digunakan lima variabel independen, yaitu tingkat pengangguran terbuka (TPT), jumlah penduduk (JP), indeks pembangunan literasi membaca (IPLM), pengeluaran per kapita (PPK), dan jumlah kecamatan (JK). Kombinasi ketiga metode ini

berpotensi meningkatkan akurasi klasifikasi dan memberikan pemahaman yang lebih mendalam tentang faktor-faktor yang memengaruhi disparitas pembangunan manusia. Selain itu, penelitian ini menawarkan pendekatan *hybrid machine learning* berbasis spasial yang tidak hanya bersifat numerik tetapi juga kontekstual, hal ini memungkinkan evaluasi yang lebih akurat dalam skala besar dan relevan bagi pengambil kebijakan dalam menyusun strategi pembangunan yang lebih tepat sasaran.

Penelitian ini bertujuan untuk mengembangkan penelitian terdahulu, yaitu klasifikasi dan analisis spasial IPM dengan *machine learning* dan GWR di Indonesia Bagian Barat. Untuk algoritma *machine learning* yang digunakan, yaitu KNN dan SVM. Dengan mengintegrasikan metode GWR, KNN, dan SVM akan mempunyai probabilitas analisis yang lebih komprehensif dalam memahami distribusi IPM dan klasifikasinya berdasarkan faktor-faktor spasial.

KAJIAN TEORI

IPM merupakan indikator komposit yang diperkenalkan oleh United Nations Development Programme untuk mengukur kualitas hidup masyarakat melalui tiga dimensi utama, yaitu kesehatan, pendidikan, dan standar hidup. IPM tidak hanya menilai pertumbuhan ekonomi, tetapi juga kesejahteraan manusia secara menyeluruh, sehingga menjadi acuan penting dalam evaluasi pembangunan. Keberhasilan pembangunan nasional tidak hanya dilihat dari ekonomi, tetapi juga dari capaian IPM. Meskipun IPM Indonesia terus meningkat, disparitas antar wilayah, khususnya di Indonesia bagian barat, masih menjadi tantangan. Oleh karena itu, diperlukan klasifikasi IPM dan analisis spasial untuk memahami faktor-faktor yang memengaruhi perbedaan tersebut agar kebijakan pembangunan lebih tepat sasaran.

Algoritma KNN

Algoritma KNN adalah metode klasifikasi terhadap data yang berlandaskan pada jarak terdekat pada data tersebut. Nilai k sangat berpengaruh pada pengklasifikasian KNN sebagai pembatas jarak terdekat yang akan digunakan. Pada penelitian ini, digunakan jarak *Manhattan* untuk mengukur jumlah bobot absolut dari perbedaan antara kasus yang sedang diuji dan kasus lain dalam basis kasus. Untuk menghitung bobot (jarak *Manhattan*), digunakan persamaan berikut:

$$d_M(x,y) = \sum_{i=0}^{n} |x_i - y_i|, \quad (1)$$

x dan y adalah dua data dalam ruang dua dimensi, dan n adalah jumlah dimensi dalam ruang tersebut. Semakin kecil nilai bobot (jarak *Manhattan*), semakin dekat data x_i dengan data y_i dalam ruang berdimensi dua, dan semakin besar nilai bobot, semakin jauh keduanya.

Algoritma SVM

Metode SVM bertujuan untuk menentukan fungsi pemisah atau klasifier yang optimal guna membedakan antara dua set data yang berbeda. SVM bekerja dengan memetakan data ke dalam ruang berdimensi tinggi dan mencari garis atau bidang pemisah (hyperplane) yang memaksimalkan jarak antara data dari dua kelas terdekat, yang disebut *support vectors*. *Hyperplane t*erbaik adalah *hyperplane* yang berada di posisi tengah antara objek dari masing-masing kelas. Adapun fungsi linier dari SVM adalah sebagai berikut:

Predict	*Actual*	
	True	*False*
True	*True Positif*	*False Negatif*
False	*False Positif*	*True Negatif*

$$g(x) = sign(f(x)) \quad (2)$$

dengan

$$f(x) = w^t x + b \quad (3)$$

$x, w \epsilon \mathbb{R}^n$ dan $b \epsilon \mathbb{R}$. Permasalahan klasifikasi ini dapat dirumuskan dengan menemukan nilai parameter (w, b) sehingga persamaan $sgn(f(x_i)) = sgn(< w, x > +b) = y_i$ terpenuhi untuk setiap i. Misalkan himpunan $X = \{x_1, x_2, \dots, x_n\}$, dinyatakan sebagai kelas positif jika $f(x) \geq 0$ serta yang lainnya termasuk kedalam negatif. SVM melakukan klasifikasi himpunan vektor training berupa set data berpasangan dari dua kelas , yaitu $(x_i, y_i), x_i \epsilon \mathbb{R}^n, y_i \epsilon \{1, -1\}, i = 1, \dots, n.$

Hyperplane yang optimal dengan menterbesarkan $\frac{2}{\|w\|}$ atau meminimalkan $\varphi(w) = \frac{1}{2}\|w\|^2$. Jika data tidak dapat dipisahkan secara linier, SVM menggunakan kernel trick untuk mengubah ruang fitur sehingga data menjadi lebih terpisahkan. Dengan pendekatan ini, SVM mampu menghasilkan model klasifikasi yang akurat dan robust terhadap outlier serta data dengan dimensi tinggi.

Evaluasi Model

Setelah proses perhitungan prediksi dilakukan pada data training menggunakan algoritma KNN dan SVM, tahapan berikutnya adalah evaluasi model untuk membandingkan nilai aktual dengan hasil prediksi. Metode evaluasi yang diterapkan dalam penelitian ini adalah *confusion matrix* (matriks

kebingungan), yang merupakan tabel yang menggambarkan kinerja model klasifikasi. *Confusion matrix* menghasilkan empat nilai yang mencerminkan hasil perhitungan dari evaluasi model.

Tabel 1. *Confussion matrix* pengklasifikasi biner

Predict	*Actual*	
	True	*False*
True	*True Positif*	*False Negatif*
False	*False Positif*	*True Negatif*

dengan menggunakan *confusion matrix*, dapat dihitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score yang memberikan wawasan lebih dalam tentang kinerja model dalam melakukan klasifikasi.

Geographically Weighted Regression

GWR merupakan pengembangan model regresi yang memperhitungkan variasi spasial dalam hubungan antar variabel. Dengan tahapan :

1. menentukan nilai bandwidth optimum dilihat dari nilai AIC atau *cross validation* (CV) menggunakan *adaptive gaussian* :

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h_i}\right)^2\right)^2 & , untuk\ d_{ij} \le h_i \\ 0 & , untuk\ lainnya \end{cases}$$

2. membuat model GWR dengan bandwidth optimum.
   Model GWR digunakan untuk mendeskripsikan keterkaitan dari masing-masing variabel terhadap IPM. Adapun bentuk umum model GWR :

$$y_j = \beta_0(u_j, v_j) + \sum_{k=1}^{p} \beta_k(u_j, v_j) x_k + \varepsilon_j$$

$$j = 1,2, \dots, J$$

Pemetaan hasil parameter model GWR di Indonesian Bagian Barat menggunakan software ArcGis yang nantinya akan terlihat pengaruh IPM dimasing-masing kabupaten/kota .

Pengujian hipotesis dalam model GWR dilakukan melalui dua pendekatan, yaitu kesesuaian model dan uji parameter .

GWR Uji Statistik Model

Pengujian ini dilakukan untuk menentukan apakah model GWR secara signifikan memberikan hasil yang lebih baik dalam memodelkan data dibandingkan model *Ordinary Least Squares* (OLS). Hipotesis yang dirumuskan adalah:

H_0 : $\beta_k(u_j, v_j) = \beta_k$ untuk masing-masing$k = 0,1,2, \dots, p$ dan $j = 1,2, \dots, J$

H_1 : ada setidaknya satu $\beta_k(u_i, v_i) \neq \beta_k$Bahasa Indonesia:$k = 0,1,2, \dots, p$

Statistik uji yang digunakan dapat ditulis sebagai berikut (Brunsdon, et.al , 1999).

$$F = \frac{RSS(H_o)/df_1}{RSS(H_1)/df_2} \quad (4)$$

Apabila nilai F cenderung kecil, menunjukkan bahwa hipotesis alternatif (H1) lebih sesuai. Sebagai tambahan, GWR menawarkan tingkat kecocokan model yang lebih unggul dibandingkan regresi global. Jika tingkat signifikansi (α) diberikan, maka keputusan diambil dengan mengurangkan (H_0) ketika $F < F_{1-\alpha,df1,df2}$ dimana $df_1 = \frac{{\delta_1}^2}{\delta_2}$ dan $df_2 = (n - p - 1)$.

Pengujian Parameter Model

Tujuan dari pengujian ini adalah untuk mengetahui paramater mana yang secara statistik berpengaruh signifikan terhadap variabel independen. Hipotesisnya dirumuskan sebagai berikut:

H_0 : $\hat{\beta}_k(u_i, v_i) = 0$ untuk setiap k = 0,1,2,...,p

H_1 : setidaknya ada satu $\hat{\beta}_k(u_j, v_j) \neq 0$

Statistik yang digunakan dalam uji ini adalah :

$$t_{hit} = \frac{\hat{\beta}_k(u_j, v_j)}{Se(\hat{\beta}_k(u_j, v_j))} \quad (6)$$

dengan $Se(\hat{\beta}_k(u_j, v_j))$ adalah standar error yang diperoleh dari akar estimasi varians parameter GWR. Kriteria pengujian yang digunakan adalah, jika $|t_{hit}| > t_{(\alpha/2, df)}$ maka keputusannya H_0 dapat ditolak atau $\hat{\beta}_k(u_j, v_j) \neq 0$.

Pemetaan hasil parameter model GWR di Indonesian Bagian Barat menggunakan software ArcGis yang nantinya akan terlihat pengaruh IPM dimasing-masing kabupaten/kota .

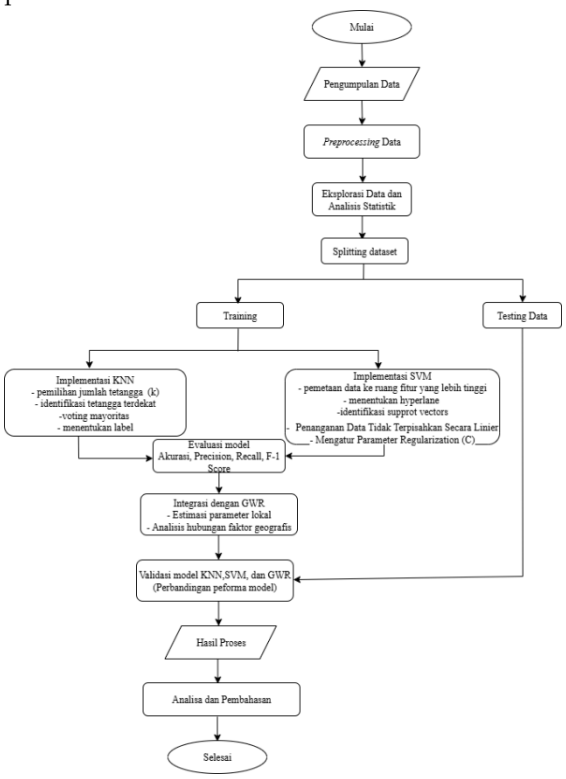
Pada tahap akhir dilakukan analisis mengenai hasil estimasi yang diperoleh. Hasil analisis ini akan menjadi rujukan bagi pemerintah dan masyarakat dalam upaya meningkatkan akurasi dan keandalannya dalam analisis spasial Indeks Pembangunan Manusia.

METODE

Data yang digunakan dalam penelitian ini merupakan data sekunder mengenai Indeks Pembangunan Manusia di Indonesia bagian barat pada tahun 2024. Data diperoleh dari situs resmi Badan Pusat Statistik (BPS) dengan total sebanyak

259 Kabupaten/Kota. Data tersebut dibagi menjadi dua bagian, yaitu data pelatihan (training) dan data pengujian (testing), dengan proporsi pembagian sebesar 70% untuk data pelatihan dan 30% untuk data pengujian. Penelitian ini, model IPM diklasifikasi dengan *machine learning*, yaitu metode KNN dan SVM.

Kemudian dilakukan analisis spasial dengan metode GWR untuk memahami faktor-faktor spasial yang memengaruhi disparitas pembangunan manusia di Indonesia bagian barat. Adapun variabel yang digunakan dalam penelitian ini, yaitu IPM, tingkat pengangguran terbuka (TPT), Pengeluaran Per-Kapita (PPK), Jumlah Penduduk (JP), Indeks Pembangunan Literasi Membaca (IPLM), dan Jumlah Kecamatan (JK). Perangkat lunak yang digunakan dalam analisis, yaitu Python, R, dan ArcGIS. Adapun diagram alir dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Tahapan dari *preprocessing* data ini, yaitu data cleaning, data integration, dan transformasi data. Apabila data pemeriksaan IPM yang digunakan sudah lengkap, tidak memiliki noise dan sudah konsisten maka dapat melanjutkan ke tahap selanjutnya.

Pada proses Eksplorasi data dan analisis statistik, akan digunakan berbagai teknik dan parameter dalam *descriptive statistics* yang bertujuan untuk menemukan pola, hubungan, serta membangun intuisi terkait data yang diolah. Selain itu, akan digunakan visualisasi data untuk menemukan pola dan memvalidasi parameter *descriptive statistics* yang diperoleh. Pada fase awal ini, dilakukan pemeriksaan deskriptif guna memahami karakteristik yang terdapat dalam data. Apabila terdapat variasi dalam data, proses normalisasi data digunakan untuk menyamakan skala dari variabel-variabel tersebut. Selanjutnya, dataset dipisahkan menjadi dua bagian, yakni *testing* dan *training*.

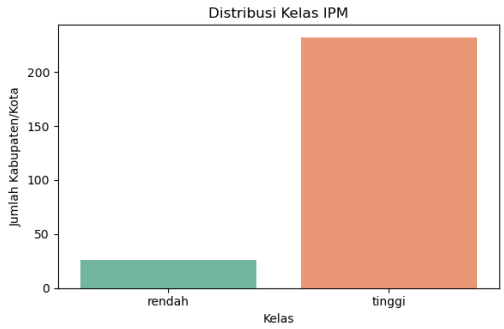
HASIL DAN PEMBAHASAN

Preprocessing dan Eksplorasi data

Dalam penelitian ini, IPM dijadikan sebagai variabel target (*variabel dependen* atau *variabel Y*) karena mencerminkan tingkat kualitas hidup penduduk pada setiap kabupaten/kota di wilayah Indonesia bagian barat. Adapun variabel independent atau variable *x* yang digunakan dalam penelitian ini, yaitu TPT (Tingkat Pengangguran Terbuka), JP (Jumlah Penduduk), IPLM (Indeks Pengeluaran Lajur Menengah), RPPK (Rata-rata Pengeluaran per Kapita), dan JK (Jumlah Kematian), Latitude, dan Longitude. Namun variable *x* utama yang dilakukan analisis, yaitu TPT, JP, IPLM, RPPK, dan JK. Hasil analisis menunjukkan bahwa seluruh variabel memiliki jumlah entri yang sama, yaitu sebanyak 258 observasi tanpa adanya nilai hilang maupun duplikasi data, sehingga data dianggap bersih dan siap untuk dianalisis lebih lanjut.

Distribusi data menunjukkan bahwa 50% dari kabupaten/kota memiliki TPT kurang dari 4,95, IPLM di bawah 60,22, dan JK tidak melebihi 13 kasus, yang menunjukkan kecenderungan bahwa sebagian besar wilayah berada pada kondisi sosial ekonomi yang relatif moderat. Namun, nilai maksimum pada variabel JP (376.489), RPPK (24.975), dan JK (666) menunjukkan adanya kabupaten/kota dengan karakteristik yang sangat ekstrem dibandingkan wilayah lainnya. Hal ini mengindikasikan pentingnya proses normalisasi data pada tahap *preprocessing* sebelum dilakukan analisis klasifikasi atau klasterisasi lanjutan.

Selanjutnya, nilai pada variabel target Kelas, yang sebelumnya direpresentasikan secara numerik (0 untuk IPM rendah dan 1 untuk IPM tinggi), diubah menjadi bentuk kategorikal ('rendah' dan 'tinggi') guna mempermudah interpretasi hasil analisis.

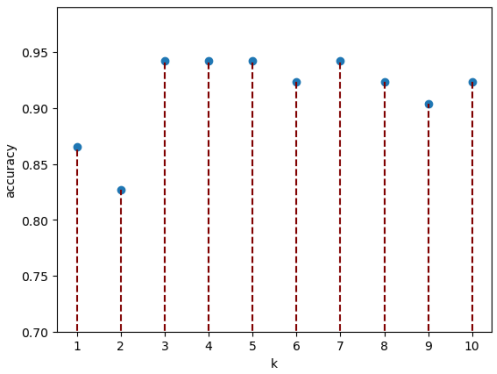


Gambar 2. Distribusi kelas IPM

Berdasarkan Gambar 2. hasil eksplorasi data menunjukkan bahwa sebagian besar wilayah di Indonesia bagian barat tergolong dalam kelas IPM "tinggi", yaitu sebanyak 232 wilayah, sedangkan hanya 26 wilayah yang termasuk dalam kelas IPM "rendah". Distribusi ini mengindikasikan adanya ketimpangan pembangunan yang masih perlu ditelaah lebih lanjut, khususnya pada wilayah dengan klasifikasi IPM rendah, agar kebijakan pembangunan dapat diarahkan secara lebih tepat sasaran.

Machine Learning

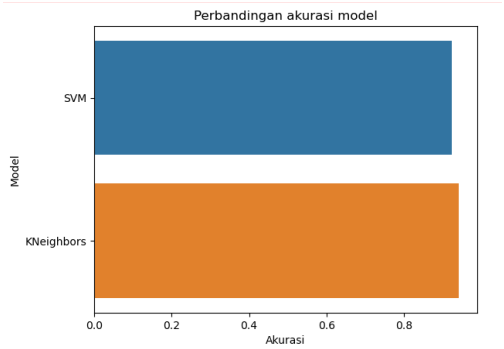
Metode *machine learning* yang digunakan dalam penelitian ini terdiri atas dua algoritma utama, yaitu KNN dan SVM. Pada algoritma KNN salah satu parameter penting yang digunakan dalam proses klasifikasi, yaitu jumlah tetangga terdekat (k). Oleh karena itu, dilakukan serangkaian eksperimen untuk mengevaluasi performa model terhadap variasi nilai k, yaitu dari k = 1 hingga k = 10. Proses evaluasi dilakukan dengan membandingkan nilai akurasi dari masing-masing model untuk setiap nilai k terhadap data uji (*testing set*).



Gambar 3. Hasil akurasi KNN pada k=1-10

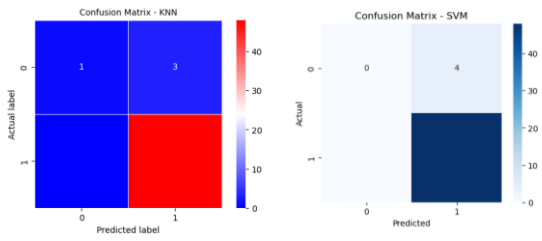
Berdasarkan Gambar 3. hasil evaluasi menunjukkan bahwa nilai akurasi cenderung meningkat seiring dengan bertambahnya nilai k, hingga mencapai nilai optimal pada k = 3,4,5 dan k = 7, dengan tingkat akurasi mencapai 94,23%. Hal ini menunjukkan bahwa pada titik tersebut model memiliki keseimbangan terbaik antara kompleksitas dan generalisasi. Berdasarkan hasil tersebut, nilai k = 5 dipilih sebagai parameter optimal dalam membangun model klasifikasi akhir .

Tahap selanjutnya dalam penelitian ini dilakukan dengan menerapkan metode SVM sebagai algoritma pembanding. Berdasarkan hasil evaluasi, diperoleh nilai akurasi sebesar 0,923 atau 92,31% pada data uji.



Gambar 4. Perbandingan akurasi model KNN dan SVM

Gambar 4. menunjukkan bahwa nilai akurasi (*test score*) yang untuk model KNN, yaitu sebesar 0,9423 atau 94,23%, sedangkan model SVM mempunyai akurasi sebesar 92,3%. Hal ini menunjukkan bahwa model KNN memiliki performa klasifikasi yang lebih baik dibandingkan model SVM dalam pengujian yang dilakukan. Meski demikian, karena distribusi kelas dalam data bersifat tidak seimbang, diperlukan evaluasi lebih lanjut menggunakan metrik tambahan seperti *confusion matrix*, *precision*, *recall*, dan *f1-score* untuk menilai kemampuan model dalam mengenali masing-masing kelas secara proporsional.



Gambar 5. Perbadingan hasil *confussion matrix* Algoritma KNN dan SVM

Berdasarkan Gambar 5. *confusion matrix* algoritma KNN menunjukkan model memprediksi dengan

benar 48 sampel kelas 1 (True Positive) dan 1 sampel kelas 0 (True Negative). Namun, model juga melakukan kesalahan klasifikasi terhadap 3 sampel kelas 0 yang diprediksi sebagai kelas 1 (False Positive), serta 0 sampel kelas 1 yang diprediksi sebagai kelas 0 (False Negative). Hal ini menunjukkan bahwa meskipun akurasi model tergolong tinggi, masih terdapat sejumlah kesalahan klasifikasi.

Sebaliknya, pada *confusion matrix* algoritma SVM, model tidak mampu mengklasifikasikan sampel kelas 0 dengan benar. Seluruh 4 sampel kelas 0 diklasifikasikan secara keliru sebagai kelas 1 (False Positive). Namun demikian, model berhasil mengklasifikasikan seluruh 48 sampel kelas 1 dengan benar (True Positive). Hal ini menunjukkan bahwa SVM memiliki kecenderungan untuk mengklasifikasikan seluruh sampel sebagai kelas 1, yang berdampak pada menurunnya performa model dalam mendeteksi kelas 0.

Tabel 2. Perbadingan hasil *precision*, *recall*, dan *f1-score*

Algoritma	*precision*	*recall*	*f1-score*
KNN	0.94	0,94	0,92
SVM	0,85	0,92	0,88

Berdasarkan Tabel 2. Nilai *precision* pada algoritma KNN diperoleh sebesar 0,94, yang mengindikasikan bahwa dari seluruh prediksi positif yang dihasilkan oleh model, sebesar 94% merupakan prediksi yang benar. Angka ini menunjukkan bahwa kesalahan klasifikasi positif palsu (*false positive*) dapat diminimalkan dengan cukup baik oleh KNN. Sementara itu, precision pada algoritma SVM tercatat lebih rendah, yaitu 0,85, yang menunjukkan bahwa proporsi prediksi positif yang benar masih di bawah algoritma KNN.

Dari segi recall, algoritma KNN kembali menunjukkan performa yang tinggi dengan nilai 0,94, yang berarti bahwa 94% dari seluruh sampel positif berhasil diidentifikasi secara benar oleh model. Sedangkan pada algoritma SVM, nilai recall yang diperoleh adalah 0,92, yang meskipun cukup tinggi, tetap menunjukkan bahwa terdapat sampel positif yang gagal teridentifikasi secara tepat.

Selanjutnya, nilai f1-score, yang merupakan rata-rata harmonis dari precision dan recall, dihitung untuk memberikan evaluasi menyeluruh terhadap kinerja model. Algoritma KNN memperoleh nilai f1-score sebesar 0,92, yang mencerminkan keseimbangan yang baik antara kemampuan model dalam mengenali kelas positif serta menghindari kesalahan klasifikasi. Sebaliknya, algoritma SVM menghasilkan f1-score sebesar 0,88, yang mengindikasikan bahwa secara keseluruhan, performa klasifikasi SVM masih berada di bawah KNN.

Dengan demikian, dapat disimpulkan bahwa algoritma KNN menunjukkan performa yang lebih baik dibandingkan algoritma SVM, berdasarkan ketiga metrik evaluasi yang digunakan. Hasil ini konsisten dengan *confusion matrix* yang telah disajikan sebelumnya.

GWR

Regresi adalah teknik yang digunakan utuk menilai sejauh mana variabel respons dipengaruhi oleh variabel prediktor. Keberhasilan estimator linear tak bias terbaik sangat dipengaruhi oleh beberapa asumsi fundamental. Model regresi yang dihasilkan dengan menggunakan metode kuadarat terkecil didasari oleh asumsi-asumsi klasik ini. Keakuratan dan validitas analisis regresi sangat bergantung pada peran penting asumsi ini. Asumsi klasik mencakup homoskedastisitas, tidak adanya autokorelasi, tidak adanya multikolineritas, serta kenormalan residual. Setelah pengujian asusmis dilakukan, dapat disimpulkan bahwa model regresi memenuhi sebagian besar syarat yang diperlukan, kecuali dalam homoskedastisitas. Berdasarkan Tingkat signifikansi (bp < 0,05) yang diperoleh dari pengujian, diputuskan untuk melanjutkan penggunaan GWR sebagai Solusi masalah heteroskedastisitas.

Pemilihan fungsi kernel merupakan aspek penting yang harus diperhatikan dalam pemodelan GWR. Fungsi kernel bertindak sebagai fungsi pembobotan yang memiliki peran krusial dalam estimasi parameter pada model GWR. Pentingnya pemilihan fungsi kernel yang tepat karena akan mempengaruhi secara langsung bobot spasial yang diterapkan pada setiap observasi dalam analisis. Semakin dekat data dengan titik regresi i, semakin besar bobot yang diberikan dibandingkan dengan data yang lebih jauh. Dalam penelitian ini, fungsi kernel *adaptive gaussian* diterapkan pada data tersebut, dengan nilai AIC yaitu 1183.08.

Estimasi parameter model GWR menggunakan metode *Weighted Least Square* (WLS). Nilai bobot tersebut merepresentasikan data observasi yang satu dengan yang lainnya. Oleh karena itu, pembobotan dalam GWR memiliki peran yang sangat penting. Adapun estimasi GWR dengan matriks pembobotan adaptif *Gaussian* dapat dilihat pada Tabel 3 .

Tabel 3. Estimasi Parameter di Provinsi Jakarta

Daerah
$Y_{\text{Kepulauan Seribu}}$ = 134736577,3 JP + 2,16 IPLM + 1,35 PPK + 2,75 TPT + -2,03 JK
$Y_{\text{Kota Jakarta Selatan}}$ = 126522974,4 JP + 2,64 IPLM + 1,35 PPK + 2,46 TPT + -3,80067 JK
$Y_{\text{Kota Jakarta Timur}}$ = 124281879,9 JP + 2,71 IPLM + 1,35 PPK + 2,39 TPT + -3,94 JK
$Y_{\text{Kota Jakarta Pusat}}$ =128134697,9 JP + 2,60 IPLM + 1,35 PPK + 2,51 TPT + -3,72 JK
$Y_{\text{Kota Jakarta Barat}}$ = 129046110 JP + 2,53 IPLM + 1,356 PPK + 2,55 TPT + -3,507 JK
$Y_{\text{Kota Jakarta Utara}}$ = 127622661,8 JP + 2,62 IPLM + 1,355 PPK + 2,49 TPT + -3,75 JK
$Y_{\text{DKI Jakarta}}$ = 126922581,8 JP + 2,63 IPLM + 1,356 PPK + 2,47 TPT + -3,78 JK

Uji Signifikansi Model GWR

Pengujian signifikansi terhadap model GWR dilakukan untuk menentukan apakah model tersebut secara signifikan lebih cocok dengan data dibandingkan model OLR. Hipotesisnya dirumuskan sebagai berikut:

$H_1 : \beta_1 = \beta_2 = \beta_3 = 0$

H_2 : Setidaknya satu tanda sama dengan tidak terjadi .

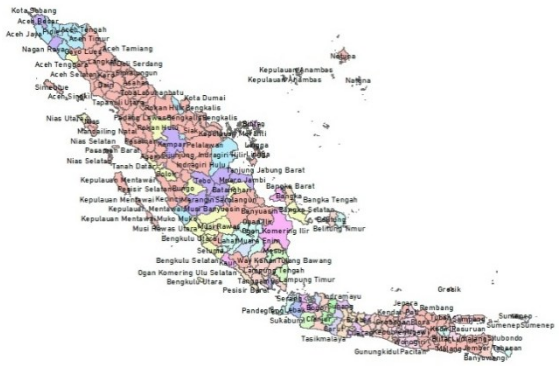
Tabel 4. ANOVA

Model	Sum of Square	Df	Mean Square	F
OLS Residuals	1490.33	6.0000		
GWR Improvement	541.93	18.42	29.4207	
GWR Residuals	948.40	264.58	3.5845	8.2077

Pada tabel 4, dengan tingkat keyakinan 5%. Hasil yang diperoleh dari perangkat lunak R untuk fungsi pembobotan adaptive gaussian ditunjukkan pada tabel 3. Hasilnya terlihat pada perhitungan p-value < 0,05 yaitu 0.0001047 dan F-hitung adalah (8,2077 > 2,11), Hal ini menunjukkan bahwa terdapat perbedaan yang signifikan secara statistik antara model regresi dan model GWR.

Pemetaan Angka Indeks Pembangunan Manusia

Tahap selanjutnya adalah memetakan hasil estimasi angka indeks Pembangunan manusia di Indonesia Bagian Barat Tahun 2023. Pemetaan hasil estimasi dengan menggunakan model GWR terhadap angka Indeks Pembangunan Manusia untuk masing-masing Kabupaten/Kota di Indonesia Bagian Barat Tahun 2023 sebagai berikut :



Gambar 6. Peta sebaran Indeks Pembangunan Manusia di tiap kabupaten/kota

Gambar 6. menjelaskan bahwa terdapat 16 klasifikasi warna pada peta tersebut. Setiap warna mewakili faktor apa saja yang mempengaruhi tingkat indeks pembangunan manusia di setiap kabupaten/kota di Indonesia Bagian Barat. Salah satunya provinsi Jakarta, dimana kepulauan seribu dipengaruhi oleh JP, TPT, dan JP. Sedangkan Kota Jakarta Selatan, Jakarta Timur, Jakarta Pusat, Jakarta Barat, Jakarta Utara, dan DKI Jakarta dipengaruhi oleh JP dan TPT.

PENUTUP

SIMPULAN

Berdasarkan hasil dan pembahasan, penelitian ini menunjukkan bahwa integrasi *Machine Learning* dengan model *Geographically Weighted Regression* (GWR) memberikan hasil yang efektif dalam

menganalisis faktor-faktor spasial yang memengaruhi IPM di Indonesia bagian barat. Dari sisi performa klasifikasi, algoritma KNN memberikan hasil akurasi tertinggi sebesar 94,23%, dibandingkan dengan SVM yang menghasilkan akurasi 92,30%.

Sementara itu, penerapan GWR dengan pembobot *adaptive Gaussian* berhasil mengelompokkan wilayah studi ke dalam 16 kelompok spasial dengan kombinasi faktor-faktor dominan yang berbeda. Faktor-faktor utama yang memengaruhi IPM meliputi jumlah penduduk (JP), indeks pembangunan literasi manusia (IPLM), pengeluaran per kapita (PPK), tingkat pengangguran terbuka (TPT), dan jumlah kecamatan (JK). Contohnya, model GWR untuk Kota Jakarta Selatan mengidentifikasi kontribusi signifikan dari keempat variabel tersebut terhadap IPM lokal. Salah satunya adalah Kota Jakarta Selatan dengan model GWR-nya adalah

$$Y_{jakarta\ selatan} = 126522974{,}4\text{JP} + 2{,}64267\text{E}+14\text{IPLM} + 1{,}35655\text{E}+14\text{PPK} - 3{,}80067\text{E}+14\text{TPT}.$$

Secara keseluruhan, integrasi pendekatan *Machine Learning* dan GWR ini tidak hanya mampu meningkatkan akurasi klasifikasi, tetapi juga memberikan wawasan spasial yang lebih mendalam terhadap faktor-faktor pembangunan manusia. Nilai koefisien determinasi (R^2) sebesar 84,17% menunjukkan bahwa model yang dihasilkan sangat andal dalam menjelaskan variabilitas IPM berdasarkan konteks lokal wilayah. Temuan ini dapat menjadi dasar untuk perumusan kebijakan pembangunan yang lebih tepat sasaran dan berbasis data spasial.

SARAN

Berdasarkan hasil penelitian, disarankan agar penelitian selanjutnya mengembangkan model dengan mencoba algoritma *Machine Learning* lain seperti Random Forest atau XGBoost untuk memperoleh performa yang lebih optimal, serta menambahkan variabel-variabel penting seperti kemiskinan, infrastruktur, dan akses layanan kesehatan agar analisis IPM lebih komprehensif. Selain itu, cakupan wilayah penelitian perlu diperluas ke seluruh Indonesia dan dikombinasikan dengan pendekatan spasial lanjutan seperti GTWR untuk menangkap dinamika ruang dan waktu secara lebih baik

DAFTAR PUSTAKA

Bishop, Christopher M., and Nasser M. Nasrabadi. 2006. *Pattern Recognition and Machine Learning*. Vol. 4. Springer.

Darsyah, Moh Yamin. 2017. "Lasifikasi Indeks Pembangunan Manusia (Ipm) Dengan Pendekatan k-Nearset Neighbor (k-Nn)." in *Prosiding Seminar Nasional \& Internasional*.

Fauzi, Fatkhurokhman. 2017. "K-Nearset Neighbor (k-Nn) Dan Support Vector Machine (Svm) Untuk Klasifikasi Indeks Pembangunan Manusia Provinsi Jawa Tengah." *Indonesian Journal of Mathematics and Natural Sciences* 40(2):118–24.

Firmansyah, Gustian Angga, Junta Zeniarja, Harun Al Azies, Syuhra Putri Ganiswari, and others. 2023. "Machine Learning-Enhanced Geographically Weighted Regression for Spatial Evaluation of Human Development Index across Western Indonesia." *Journal of Applied Geospatial Information* 7(2):996–1003.

Indonesia, Badan Pusat Statistik Republik. 2017. "Indeks Pembangunan Manusia 2016." *Jakarta (ID): Badan Pusat Statistik*.

Jun, Zhao. 2021. "The Development and Application of Support Vector Machine." P. 52006 in *Journal of Physics: Conference Series*. Vol. 1748.

Krismayanto, Ujang Kurnia, and Ernawati Pasaribu. 2022. "Analisis Regresi Spasial Indeks Pembangunan Kesehatan Masyarakat Dan Paradoks Simpson Kabupaten/Kota Di Pulau Sumatera Tahun 2018." Pp. 1037–52 in *Seminar Nasional Official Statistics*. Vol. 2022.

Latuconsina, Fidyah, Marlon Stivo Noya van Delsen, and others. 2024. "Klasifikasi Menggunakan Metode Support Vector Machine (SVM) Multiclass Pada Data Indeks Desa Membangun (IDM) Di Provinsi Maluku." *Journal of Mathematics, Computations and Statistics* 7(2):380–95.

Marizal, Muhammad, and Hayatul Atiqah. 2022. "Pemodelan Indeks Pembangunan Manusia Di Indonesia Dengan Geographically Weighted Regression (GWR)." *Jurnal Sains Matematika Dan Statistika* 8(2):133–45.

Maulana, Akbar, Renny Meilawati, and Vita Widiastuti. 2019. "Pemodelan Indeks Pembangunan Manusia (IPM) Metode Baru Menurut Provinsi Tahun 2015 Menggunakan Geographically Weighted Regression (GWR)." *Indonesian Journal of Applied Statistics* 2(1):21–33.

Oktaviana, Oktaviana, Riri Syafitri Lubis, and Rina Widyasari. 2024. "Penerapan Metode Ensemble K-Nearest Neighbor Untuk Memprediksi Indeks Pembangunan Manusia Di Provinsi Sumatera Utara." *Justek: Jurnal Sains Dan Teknologi* 7(4):334–44.

Pamungkas, Canggih Ajika, and Wahyu Wijaya Widiyanto. 2022. "Klasifikasi Indeks Pembangunan Manusia Di Indonesia Tahun 2022 Dengan Support Vector Machine." *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer* 2(3):139–45.

Safitri, Ira, Ardika Satria, and Rizty Maulida Badri. 2024. "Implementasi Algoritma K-Nearest Neighbor Dalam Klasifikasi Indeks Pembangunan Manusia Provinsi Sumatera Selatan Tahun 2023." *Digital Transformation Technology* 4(2):768–75.

Witata, Ganes Angga, and Joko Triloka. 2023. "Kajian Perbandingan Algoritma KNN Dan SVM Untuk Prediksi Pengangguran Di Provinsi Lampung." Pp. 218–23 in *Prosiding Seminar Nasional Darmajaya*. Vol. 1.

Yang, Xiaoyue, Yi Yang, Shenghua Xu, Jiakuan Han, Zhengyuan Chai, and Gang Yang. 2023. "A New Algorithm for Large-Scale Geographically Weighted Regression with k-Nearest Neighbors." *ISPRS International Journal of Geo-Information* 12(7):295.